SELF-CALIBRATING SURROUND SOUND SYSTEM

Cross-reference to other Patent Applications

This application claims the benefit of U.S. provisional Patent Application No.
60/198,927, filed 04/21/2000, which is incorporated herein by reference in its entirety.

5   Field of the Invention

The invention is directed to a multi-channel surround sound system, and more
particularly to a surround sound system allowing automatic calibration and adjustment
of the frequency, amplitude and time response of each channel.

Background of the Invention

10   "Surround sound" is a term used in audio engineering to refer to sound reproduction
systems that use multiple channels and speakers to provide a listener positioned between
the speakers with a simulated placement of sound sources. Sound can be reproduced
with a different delay and at different intensities through one or more of the speakers to
"surround" the listener with sound sources and thereby create a more interesting or
15   realistic listening experience.

Multi-channel surround sound is employed in movie theater and home theater
applications. In one common configuration, the listener in a home theater is surrounded
by five speakers instead of the two speakers used in traditional home stereo system. Of
the five speakers, three are placed in the front of the room, with the remaining two
20   surround speakers located to the rear or sides (THX dipolar) of the listening/viewing

position. Among the various surround sound formats in use today, Dolby® Surround™ is the original surround format, developed in the early 1970's for movie theaters. Dolby® Digital™ made its debut in 1996 and is installed in more than 30,000 movie theaters and 31 million home-theater products. Dolby Digital is a digital format with six

5   discrete audio channels and overcomes certain limitations of Dolby Surround which relies on a matrix system that combines four audio channels into two channels to be stored on the recording media. Dolby Digital is also called a 5.1-channel format and was universally adopted several years ago for film-sound recording. Yet another new format is called Digital Theater System (DTS). DTS offers higher audio quality than Dolby

10   Digital (1,411,200 versus 384,000 bits per second) as well as an optional 7.1 configuration.

The audio/video preamplifier (or A/V controller) handles the job of decoding the two-channel Dolby Surround, Dolby Digital, or DTS encoded signal into the respective separate channels. The A/V preamplifier output provides six line level signals for the

15   left, center, right, left surround, right surround, and subwoofer channels, respectively. These separate outputs are fed to a multiple-channel power amplifier or as is the case with an integrated receiver, are internally amplified, to drive the home-theater loudspeaker system.

Manually setting up and fine-tuning the A/V preamplifier for best performance can be

20   demanding. After connecting a home-theater system according to the owners' manuals, the preamplifier or receiver for the loudspeaker setup have to be configured. For

example, the A/V receiver or preamplifier must know the loudspeaker type, so that the bass can be directed appropriately. For example, receivers may classify loudspeakers as "large" or "small". Selecting a "small" loudspeaker will keep low-bass signals out of the speaker. This configuration is used when a subwoofer is used to reproduce low bass

5   instead of the left and right speakers. If the system has no subwoofer and full-range left and right speakers, a "large" speaker setting should be selected. The setup may also require selecting "small" or "large" surround speakers. Next a center channel speaker mode ("normal" or "wide") needs to be selected, as well as an appropriate center-channel delay so that the sound from all three front speakers arrives at a listener's ear at

10   the same time. An additional short delay for the signal to the surround speakers of typically 20ms may also have to be set to improve the apparent separation between front and rear sound.

In addition, the loudness of each of the audio channels (the actual number of channels being determined by the specific surround sound format in use) should be individually

15   set to provide an overall balance in the volume from the loudspeakers. This process begins by producing a "test signal" in the form of noise sequentially from each speaker and adjusting the volume of each speaker independently at the listening/viewing position. The recommended tool for this task is the Sound Pressure Level (SPL) meter. This provides compensation for different loudspeaker sensitivities, listening-room

20   acoustics, and loudspeaker placements. Other factors, such as an asymmetric listening space and/or angled viewing area, windows, archways and sloped ceilings, can make

calibration much more complicated

It would therefore be desirable to provide a system and process that automatically calibrates a multiple channel sound system by adjusting the frequency response, amplitude response and time response of each audio channel. It is moreover desirable

5 that the process can be performed during the normal operation of the surround sound system without disturbing the listener.

Summary of the Invention

The invention is directed to a surround sound system with an automatic calibration feature for adjusting audio channel responses to the characteristic of the listening

10 environment. The invention is also directed to a method that provides calibration and adjustment of the frequency, amplitude and time response of each channel of the surround sound system in a manner that is unobtrusive to a listener and can be employed during the listening experience of the listener.

According to one aspect of the invention, an auto-calibrating surround sound (ACSS)

15 system includes an electro-acoustic converter, such as a loudspeaker, disposed in an audio channel and adapted to emit a sound signal in response to an electric input signal. The ACSS system further includes a processor that generates a test signal represented by a temporal maximum length sequence (MLS) and supplies the test signal as part of the electric input signal to the electro-acoustic converter, and an acousto-electric converter,

20 such as a microphone, that receives the sound signal in a listening environment and

supplies a received electric signal to the processor. The processor correlates the received electric signal with the test signal in the time domain and determines from the correlated signals a whitened response of the audio channel in the listening environment.

The processor may include an impulse modeler that produces a error fit, for example, a

5 polynomial least-mean-square (LMS) fit, between a desired whitened response and the whitened response determined from the correlated signals, as well as a coefficient extractor which generates from the correlated signals filter coefficients of a corrective filter to produce the whitened response of the audio channel. The corrective filter may be located in an audio signal path between an audio signal line input and the electro-

10 acoustic converter and cascaded with the audio signal line input. The correlator and/or the IM and/or the corrective filter may be part of the processor. The processor can be a digital signal processor (DSP), and the ACSS system can further include A/D and D/A converters to enable digital processing of analog signals in the DSP.

According to another aspect of the invention, a digital filter for whitening an audio

15 channel in a listening environment includes an input receiving a digital audio signal, and a corrective filter having filter coefficients that are determined in the listening environment using a maximum length sequence (MLS) test signal. The corrective filter convolves the filter coefficients with the digital audio signal to form a corrected audio signal. An output supplies the corrected audio signal to a sound generator.

20 According to yet another aspect of the invention, a method of auto-calibrating a

surround sound system includes the acts of producing an electric calibration signal which is a maximum length sequence (MLS) signal; supplying the calibration signal to an electro-acoustic converter which converts the calibration signal to an acoustic response; and transmitting the acoustic response as a sound wave in a listening

5   environment to an acousto-electric converter. The acousto-electric converter converts the acoustic response into an electric response signal. The method further includes correlating the electric response signal with the electric calibration signal to compute filter coefficients, and cascading the filter coefficients with a predetermined channel response of the electro-acoustic converter to produce a whitened system response.

10   According to still another aspect of the invention, method of producing a matched filter for whitening an audio channel in a listening environment includes producing in the audio channel a test output sound corresponding to a temporal maximum length sequence (MLS) signal; receiving the test output sound at a predetermined location in the listening environment, thereby producing an impulse response; analyzing a

15   correlation between the impulse response and the MLS signal; and generating from the analyzed correlation filter coefficients of the matched filter.

Embodiments of the invention may include one or more of the following features. The calibration signal has a noise characteristic that is non-offensive to a listener located in the listening environment and a duration of less than approximately 3 seconds. The

20   surround sound system may include a plurality of audio channels, with each channel having at least one electro-acoustic converter, wherein the whitened response is

produced independently for each audio channel. The filter coefficients may be generated

by optimizing a "closeness of fit", for example, a least sum of squares error value,

between the polynomial model and the matched filter. Optimization of the "closeness of

fit" may include adjusting the length of the MLS signal. To produce the whitened audio

5   channel, the matched filter can be cascaded with a useful audio signal.

Further features and advantages of the present invention will be apparent from the

following description of preferred embodiments and from the claims.

Brief Description of the Drawings

The following figures depict certain illustrative embodiments of the invention in which

10   like reference numerals refer to like elements. These depicted embodiments are to be

understood as illustrative of the invention and not as limiting in any way.

Fig. 1   shows a schematic block diagram of an ACSS System;

Fig. 2   shows schematically a calibration process for the ACSS;

Fig. 3   shows the ACSS system in its operational phase;

15   Figs. 4a-b   show an uncorrected (a) and a whitened (b) frequency response of an

exemplary ACSS System;

Fig. 5   shows an exemplary minimum length sequence (MLS);

Fig. 6   shows a digital implementation of a matched moving average (FIR) filter;

Fig. 7   schematically depicts the process of whitening a channel;

Figs. 8a-b   show a simulated channel impulse response (a) and frequency response (b);

Figs. 9a-b   show the frequency response magnitude for the simulated channel impulse response of Fig. 8(a): AR model (a) and matched filter (b), both with M=5;

Figs. 10a-d   show the whitened power spectral density (PSD) for different values M of

5   the filter order: M=5 (a), M=10 (b), M=20 (c), and M=100 (d);

Fig. 11   shows a schematic block diagram of interconnected devices of the ACSS system;

Figs. 12a-b   show a satellite loudspeaker impulse response (a) and an overlay of corresponding frequency responses in an open environment (b);

10   Fig. 13   shows the frequency response of four satellite loudspeakers (a) – (d) in a listening environment; and

Figs. 14a-b   show an overlay of the original frequency response of the front-right loudspeaker (Fig. 13(b)) and simulated white frequency responses for filter order M=10 and M=50 (a) and the corresponding LMS error curve (b).

15   Detailed Description of Certain Illustrated Embodiments

The invention is directed to an auto-calibrating surround sound system that automatically adjusts the frequency response, amplitude response and time response of each audio channel without intervention from the listener. In particular, the system and method described herein can be used to whiten the frequency response of the sound

system even in changing listening environments. A signal is defined as "white" if the signal exhibits equal energy per Hz bandwidth. Accordingly, a white or whitened response of an audio system is defined as a sound output signal produced by an electro-acoustic converter, such as a loudspeaker, that exhibits equal output energy per Hz

5   bandwidth for an electric input signal to the system with equal electric energy per Hz bandwidth.

Referring first to Fig. 1, an auto-calibrating surround sound (ACSS) system 10 includes a surround sound preamplifier 12 receiving audio input signal from various conventional audio devices (not shown), such as tuners, CD and DVD players, and other digital or

10   analog signal sources, a multi-channel power amplifier 14 inserted in the signal path between the preamplifier 12 and a plurality of loudspeakers 15, 16, 17, 18, 19 located in the listening environment. The location of the loudspeakers is selected so that a listener has the impression of being surrounded by sound by, for example, placing loudspeakers 15 and 19 to the left and right behind the listener and loudspeakers 16 and 18 to the left

15   and right in front of the listener. Loudspeaker 17 is typically located at the center to covey, for example, dialog from actors shown on a TV screen. The components 12, 14 and the loudspeakers 15,.. , 19 are part of a conventional surround sound system.

As part of the auto-calibration feature, an auto-calibrating surround sound processor 13 is typically connected between the line level outputs of the preamplifier 12 and the line

20   level inputs of the multi-channel power amplifier 14. The auto-calibrating surround sound processor 13 has an additional input for a calibration microphone 11 as well as a

user control (or menu item) for initiating a calibration sequence (not shown). Once the system 10 is calibrated, the calibration microphone 11 is no longer needed and may be disconnected until the user decides to recalibrate the system.

Referring now to Figs. 2 and 3, two operating phases of the ACSS should be

5   distinguished: the calibration phase (Fig. 2) and the operational phase (Fig. 3). During the calibration phase depicted in Fig. 2, the ACSS system 20 generates a calibration signal which can be a separate signal for each loudspeaker 15, ..., in the system (the actual number of loudspeakers being determined by the desired number of channels). Typically, the center loudspeaker 17 need not be calibrated. The calibration signal is a

10   non-offensive noise, similar to white noise, which is only audible for a small amount of time (a total duration of 2-3 seconds or less). The calibration microphone 11 placed at the listener location collects the response from the loudspeakers15, ..., 19.

The calibration noise signal in the described embodiment is pseudo-random in nature and derived from a maximal length sequence (MLS) generated by MLS generator 21.

15   The signal generated by MLS generator 21 is supplied to the power amplifier 14 to drive the loudspeakers 15, ..., 19. The MLS is deterministic so that the samples received from the microphone 11 and optionally amplified in microphone preamplifier 23 can be correlated in correlator 24 with an exact replica of the MLS signal used to drive the loudspeakers, as indicated by a connection between correlator 24 and MLS generator 21.

20   The output of correlator 24 is supplied to impulse modeler 25 to derive the impulse response for a channel in the surround sound system 10. From this impulse response,

the time of flight between the listener and each loudspeaker and the frequency response

of the channel is determined. The power spectrum of the received signal is a function of

the frequency response of the power amplifier, the loudspeakers, room acoustics, and the

calibration microphone. In most cases, the dominant factors in determining the

5   frequency response is the frequency response of the loudspeakers and the room

acoustics. If any of these elements are changed or repositioned, then the power spectrum

and times of flight may change.

The measured impulse response derived from the correlator 24 is typically not well-

behaved in a mathematical sense because it is not a continuous function and therefore

10   may contain discontinuities. Some of the difficulties associated with these

discontinuities can be eliminated by forming a model of the measured impulse response.

This is done in the impulse modeler 25, which creates a recursive estimator of the

impulse response, using, for example, an auto-regressive (AR) curve fitting technique

with a polynomial model to create a least-mean-square (LMS) error curve fit to the

15   measured impulse response. This model of the impulse response is then used by

coefficient extractor 26 to generate the coefficients 27 for a matched filter to correct the

channel response.

Fig. 3 illustrates the operational phase of the ACSS system 30. Once the required filter

coefficients 27 are determined, a real-time corrective filter 32 is initialized with the

20   proper correction coefficients in the time domain for each channel in the surround sound

system. In this system, each set of coefficients defines a filter that is unique to the

requirements of the respective channel. The corrective filter 32 is placed in the audio

signal path between the surround sound preamplifier 12 and the multi-channel power

amplifier 14 to whiten the system response, as will be described in detail below. It

should be noted that the corrective filter 32 can be part of the ACSS processor 13 of Fig.

5   1. It is also possible to switch the corrective filter 32 in and out of the signal path as

needed. In addition, it should be noted that the audio signal could be either an analog, a

digital signal or some combination of analog and/or digital signals.

Fig. 4 shows the result obtained by applying the ACSS process to an exemplary low-cost

surround sound system of a type designed for personal computer systems. The top graph

10   (a) shows the uncorrected amplitude response of the system in the frequency domain.

The frequency range is limited to an upper frequency of approximately 6.5 kHz due to

the limited sampling rate of the A/D converter used to sample the original impulse

response. The lower limit of the frequency range starts at 100 Hz since the speaker is

used as a satellite speaker and hence performs poorly in reproducing low frequencies.

15   As seen in Fig. 4 (a), this particular loudspeaker has wide amplitude excursions in

excess of 20dB over the entire illustrated frequency range. Further, speaker has a

noticeable 15 dB null at approximately 2.5kHz. The bottom curve (b) shows the

frequency response of the system after ACSS correction. The majority of the previously

uncorrected amplitude excursions are now well controlled to within approximately ±2dB

20   of the nominal response. Moreover, the effect of the deep null in the original response,

although still noticeable, is significantly reduced.

The operation of the ACSS system will now be described in detail. As known from mathematical concepts, a frequency response of a system (the changes in magnitude and delay that the system imparts to sine waves of different frequencies applied to its input) has a one-to-one relationship to an impulse response (the waveform with which a system

5   responds to a sharp impulse applied to its input). The two responses can be converted into each other by a Fourier Transform and inverse Fourier Transform, respectively. Consequently, a system, such as a loudspeaker, can be characterized either by applying sine waves to find the frequency response, or by applying impulse stimuli to obtain the impulse response. Once either type of data is obtained, transformation from one to the

10   other is a simple matter of processing the Fourier transforms (typically using a computer). A narrow pulse is attractive as a measurement stimulus for several reasons. It is easy to generate using inexpensive circuitry. Both the phase and magnitude of the frequency spectrum of a narrow pulse are essentially uniform over a wide range of frequencies, allowing simultaneous measurements over most or all of the amplitude and

15   frequency ranges of a speaker and/or amplifier. Echoes in a system pulse response are easily identified and removed, so that measurements equivalent to those from an anechoic chamber can be obtained.

Since the energy of a single pulse may be small and cannot be easily increased without "clipping" in the amplifier circuitry and/or driving the loudspeaker into nonlinear

20   operation, a number of measures can be taken to increase the average power of the test signal. For example, repetitive pulse stimuli can be applied; however, to increase the

noise rejection by 30 dB, over one thousand responses may be required, resulting in an unacceptably long calibration time. Alternatively, a frequency sweep or "chirp", or so-called "pink" noise, which has an even distribution of power if the frequency is mapped in a logarithmic scale, can be employed. A full response measurement also takes a rather

5   long time, as each frequency is essentially measured separately.

A very convenient stimulus is pseudo-random noise, which is the frequency-domain version of a digital signal in the time domain known as a Pseudo-random Number (PN) pattern or Maximum Length Sequence (MLS). The magnitude of a pseudo-random noise spectrum in the frequency domain is basically flat, while the phase is scrambled - but not

10   really random. Since the spectrum is deterministic and repeatable, only a single measurement channel is required for characterizing the system.

The MLS additionally has the property that its autocorrelation function represents an impulse signal, whereas the cross-correlation function between the response of a system to an MLS with the MLS itself is the impulse response of the system which can be

15   transformed to provide the frequency response of the system, or analyzed in the time domain.

Fig. 5 illustrates an exemplary MLS of length 7, modified so that a digital "0" is represented as "-1". If a copy of the sequence is lined up exactly underneath the original sequence (autocorrelation), as indicated in the upper portion of Fig. 5, and the

20   corresponding values are multiplied and all the products are summed, a value 7 equal to

the length of the MLS is obtained. If the second sequence is shifted from the original

sequence by, for example, 5 time intervals or clock cycles, as indicated in the lower

portion of Fig. 5, which is equivalent to a time shift of an MLS signal, then the sum of

the products in this example yields a value of -1. In other words, the correlation function

5   between an N-point MLS has a sharp peak when the MLS line up exactly, with the

signal being negligibly small if an MLS response signal is misregistered with respect to

the original MLS signal. This is the underlying concept behind the ACSS system and

process.

Referring back to Fig. 2, during the calibration phase, the ACSS generates a calibration

10   signal separately for each loudspeaker in the system. Although the MLS was described

above as a sequence of δ-shaped (infinitely short) pulses, in practice an analog MLS

may have to be generated from the digital MLS, for example, by using a zero-order-hold

(ZOH) with reconstruction filter, so that the letter "S" in MLS then denotes "Signal"

rather than "Sequence."

15   As mentioned above, the system can be modeled either in the time domain or in the

frequency domain by applying a DTFT to the impulse response. In the following, the

impulse response is modeled in the time domain.

In a linear time-invariant system (LTI), a response depends on a weighted average of the

current and past M inputs $x[i]$ well as a weighted average of the most recent N outputs

20   $y[k]$:

$$y(n) = -\sum_{k=1}^{N} a_k y[n-k] + \sum_{k=0}^{M} b_k x[n-k] \qquad (1)$$

This system is sometimes also called to an Auto Regressive Moving Average (ARMA) system. An auto regressive (AR) process of order N can be described in terms of the inner product between a set of coefficients and the previous output values y[n]:

5 $$y[n] + a_1 y[n-1] + \cdots + a_N y[n-N] = v[n] \qquad (2)$$

where a_n are constant coefficients and v[n] is a white noise process used to model an error term. Since the number of coefficients will have practical limits, the impulse response may be truncated, which is equivalent to applying a window function. By recognizing that equation (2) is the convolution of the coefficients a_n and the vector

10 {y[1], ..., y[n]} of past output samples and recalling that the convolution of two time sequences can be represented as the product of their corresponding Z transforms, one obtains

$$Y(z)H_a(z) = V(z) \qquad (3)$$

where $H_a(z)$ is the Z transform of the coefficients a_n. The equation (3) shows that

15 for some process Y(z) there will be some system function H(z) that will yield the white noise process V(z).

One of the tasks in the present analysis is the determination of the transfer function H(z) for two aspects of the problem, namely to *generate* the process and to *analyze* the process. Creating a stable inverse filter is the main motivation for selecting

the model to be of type Infinite Impulse Response (IIR). In an IIR-model, the order N of the AR process in equation (2) goes to ∞. The frequency response of a linear time-invariant (LTI) system can be determined entirely in terms of its magnitude and phase $H(e^{j\omega}) = |H(\omega)|e^{j\theta(\omega)}$ by evaluating its Z transform on the unit circle, providing that the Fourier transform exists. Complications may arise from the fact that the system is not truly minimum phase, but this error will be small for typical room impulse responses.

Having selected the AR model for the system being measured, an inverse of this model is created so that the effects of the room response can be removed. Because the model is defined to be minimum-phase and stable, it will have an inverse function that is minimum phase as well. Recalling from system theory that the impulse response of cascaded stages is the convolution of the individual impulse responses of the various stages, the output sequence is as follows:

$$y[n] = \{x[n]*h_1[n]\}*h_2[n] = x[n]*\{h_1[n]*h_2[n]\} \qquad (4)$$

where x[n] is the input signal and $h_i[n]$ of the impulse responses of an individual stage i.

The next objective is to converge on an optimal set of finite impulse response (FIR) coefficients b_n for the process analyzer that will remove the effects of the room

$$y[n] = \sum_{k=0}^{M} b_k x[n-k] \qquad (5)$$

Before any coefficients can be estimated, a figure of merit may be defined so that

the performance of the model can be analyzed. This figure of merit could be the least sum of squares error between the desired matched filter output and the output of a moving average filter. In this case, if d[n] is the desired response of the matched filter, the following error ε[n] results

$$\varepsilon[n] = d[n] - \sum_{k=0}^{M} b_k h[n-k] \qquad (6)$$

Minimizing a global error term, which is computed from the sum of squared error terms γ, is done by taking the first partial derivative of γ with respect to the coefficients b_k and setting the result to zero, i.e., $\dfrac{\partial \gamma}{\partial b_k} = 0$, to find the minimum point.

This leads to a set of linear equations in terms of the cross and autocorrelation as follows

$$R_{hd}[l] = \sum_{k=0}^{M} b_k R_{hh}[l-k] \qquad (7)$$

The moving average filter that uses the coefficients b_k of equation (7) produces minimum error in the least square sense, which is the figure of merit to be optimized. This filter is also known as a Wiener-Filter and is illustrated in Fig. 6. Equation (7) can be seen as the linear convolution between the coefficients b_n and the cross correlation of the matched filter impulse response h[n].

Since the desired power spectral density (PSD) of the combined system under test (SUT) and matched filter should be flat, it can be seen that the cross correlation between d[n] and h[n] will be zero for all values of shift except at the origin, so that

equation (7) can be expressed in matrix form as

$$
\begin{bmatrix}
r_{hh}(0) & r_{hh}(1) & r_{hh}(2) & \ldots & r_{hh}(M) \\
r_{hh}(1) & r_{hh}(0) & r_{hh}(1) & \ldots & r_{hh}(M-1) \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
r_{hh}(M) & r_{hh}(M-1) & r_{hh}(M-2) & \ldots & r_{hh}(0)
\end{bmatrix}
\begin{bmatrix}
b_0 \\
b_1 \\
\vdots \\
\vdots \\
b_M
\end{bmatrix}
=
\begin{bmatrix}
h(0) \\
0 \\
\vdots \\
\vdots \\
0
\end{bmatrix}
\tag{8}
$$

As seen from the above, the minimized error term is a function not only of the coefficients b_n, but also of the filter length M. The filter length M can be selected by

5   experimental means. However, as part of automating the process, it should also be possible to select the order in an adaptive fashion, without visual inspection.

Fig. 7 is a schematic process flow diagram of an auto-calibrating process 70 that produces a whitened system response. The system monitors an input 71, for example, a signal received by calibration microphone 11. If an impulse signal is detected at 72, an

10   auto-regressive (AR) model is created using equations (1) – (3). A matched filter is created by process 75 using equations (5) – (6) and cascaded with the original channel, as described with reference to equations (4) and (7) – (8). If a global minimum error term is attained, step 77, then the system response has been optimally whitened and the auto-calibration, at least for the loudspeaker under test, is terminated in 78. Otherwise,

15   the AR model is revised in 73, possibly using a different model order determined by process step 74.

Referring now to Fig. 8a, an exemplary simulated channel impulse has the form

of an exponentially decaying sinusoidal signal that can be used to the test the

deconvolution properties of an MLS. Fig. 8b shows the corresponding frequency

response, with the spike in the frequency response corresponding to the frequency of the

dampened sinusoid. For the simulations, a model order M between M = 5 and M=100

5   was selected. The AR (auto regressive) model parameters, i.e., the filter taps of Fig. 6,

are generated as described above with reference to equations (7) and (8). The frequency

response magnitude of the AR model with M=5 is shown in Fig. 9(a). The

corresponding matched filter frequency response is shown in Fig. 9(b) and is essentially

an "inverted" AR response, i.e., the filter response has poles where the AR response has

10   zeros, and vice versa. A matched filter with a higher order of M, for example M=20,

tends to have a sharper frequency response. Finally, the matched filter of Fig. 9(b) is

cascaded with the original channel to "whiten" the channel, as seen from the process

flow of Fig. 7. Filtering the original impulse response using the matched filter should

produce an even distribution of spectral power.

15   Figs. 10(a) – (d) show the whitened power spectral density (PSD) for different

values M of the filter order between M=5 and M=100. It should be noted that the PSD is

not normalized. A filter order of M=10 or M=20 has been found to sufficiently whiten

the system response.

It should also be noted that in spite of the matched filter, a peak exclusion of 10

20   dB or more remain. The inability to reduce the peak magnitude component of this

simulation does not indicate failure of the matched filter; rather, it indicates that a lower

bound is reached. This is not considered to be a problem since most listening environments require small corrections over a wide range of frequenciesrather than the correction of a single large frequency anomaly.

Referring now to Fig. 11, the hardware of the auto calibrating surround sound
5   (ACSS) system can be implemented with standard audio components and digital signal processors. In the exemplary block diagram 110 of the ACSS of Fig. 10, the evaluation board 114 is implemented as an embedded Digital Signal Processor (DSP) 116 with onboard D/A 117 and A/D 115 converters (Texas Instruments TMS320C54x DSKplus board with C542 processor) and a 10 MHz clock. The board 114 receives suitable input
10   signals, either in digital or analog, from input device(s) 112. The other components correspond to those described above with reference to Fig. 2. Although this device has an input/output cutoff frequency significantly below 20 kHz with a 44 kHz sampling rate, it is adequate to demonstrate the validity of the proposed calibration concept. There are many other processors known in the art which can be used. Such processors, when
15   combined with higher resolution D/A and A/D converters and higher sampling rates will result in improved system performance.

As an embedded system device, the first step is to initialize the processor and corresponding peripherals. Before any of the peripherals that are included either on the C542 itself or on the DSKplus board can be used, they must be brought to the proper
20   configuration state. For example, the input ports, the filter parameters of the board's analog interface circuit (CODEC), the analog-to-digital and digital-to-analog conversion

rates are configured, and an interrupt vector table is loaded

A system under test (SUT), in this case a free space listening environment, is excited with an MLS using a loudspeaker, and a received signal is taken as the sampled output of a microphone located in the same space. The impulse response of the path

5   between the two can be deconvolved by cross-correlating the stimulus MLS with the received the signal. This is done, as described above with reference to the exemplary MLS of Fig. 5, by shifting the content of a serial port transmit register (TDXR) into the CODEC and then shifting data from the A/D converter into the serial port receive register (TRCV) and periodically convolve these data to establish the correct time scale

10   of the received signal.

An actual auto-calibration of an exemplary N-channel surround sound system is performed using four Klipsch Pro-Media v.2-400 speakers. The subwoofer and center speaker, which are typically also part of a surround sound system, are not calibrated. Each of the speakers is calibrated separately and the corresponding coefficients are

15   placed in a respective DSP memory. For performing the listening test, the matched filters can be turned on and off.

Referring now to Fig. 12, before running the four-channel surround sound test, the impulse response for each of the satellite speakers in an open laboratory space is deconvolved using the MLS technique. The system is set up so that the four frequency

20   responses can be compared. However, these measurements are not directly compared to

those that are taken in the listening environment, since the microphone placement, sound pressure level at the microphone, and the surrounding acoustic impedances can all be different. Because all four responses are similar, they are plotted in an overlay fashion. Fig 12(a) shows the impulse response of an exemplary satellite speaker (in this case, the front-right speaker in the listening environment), as well as the four overlaid frequency response magnitudes. The time of flight delay of approximately 2.2 ms indicates that the distance between the microphone and the speaker in this test was approximately 70 cm. Verifying distances like speaker placement using the exponentially determined time of flight is a good way to determine if the periodic cross-correlation is extracting the correct time base. The response feature arriving with a delay of approximately 4.3 ms indicates a first reflected signal. The sharp drop in frequency response at about 3 kHz will be the most difficult portion of the spectral response to whiten.

With the open space frequency response of each satellite speaker determined, the surround sound calibration in the actual listening environment is performed. Each of satellite speakers is calibrated individually, since even though they all have similar responses in the open space, the different placement of each speaker in the listening environment can cause the acoustic impedance to be different. Figs. 13(a) – (d) show the responses from the four loudspeakers. It should be noted that the respective pairs front-left/rear-left loudspeakers (Figs. 13(a) and 13(c)) and the front-right/rear-right loudspeakers (Figs. 13(b) and 13(d)) have a similar response, which is due to the fact that the left satellites have a rigid wall on one side, which is essentially an infinite baffle,

whereas the right satellites have no wall directly adjacent, providing a more absorbent surrounding.

Referring now to Fig. 14, the original frequency response of the front left satellite speaker was whitened using the process and system of the invention described above to illustrate that the process is capable of performing in a real listening environment. Fig. 14(a) is an overlay of the unfiltered frequency response of the front-right loudspeaker (Fig. 13(b)) and simulated whitened responses computed for filter orders M = 5 and M = 50. Fig. 14(b) shows the LMS error curve with the marked simulated orders.

While the process for automatic calibration of a surround sound system has been disclosed in connection with the preferred embodiments shown and described in detail, various modifications and improvements thereon will become readily apparent to those skilled in the art. For example, it may be desirable to differentiate between the actual impulse response information and the system noise, since it is of no interest to try and model any portion of the impulse response that is buried in the noise floor of the system. Accordingly, the results may be improved by comparing the energy, rather than the amplitude of the information carrying data which could result in an increase of the signal-to-noise ratio.

Reflections of the sound produced by a loudspeaker may also be of interest. The greater the time of flight (i.e., delay), the more phase compensation must be introduced

by the matched filter. The more severe the reflections included in the analysis, the less the system becomes the minimum phase. Minimizing the summed square error terms (LMS) to generate the coefficients for the matched filter also works best for minimum phase systems. However, with LMS, the error performance deteriorates if the system

5   becomes non-minimum phase. Systems that employ, for example, two compensation filters could be used for whitening mixed phase systems.

Because the human ear does not have a flat frequency response, a listening environment with a flat response is not necessarily the best choice. For example, an additional equalization could be added to obtain a desired preprogrammed frequency

10   response curve. In addition, since the time of flight from each loudspeaker can be determined from the measured impulse response, one skilled in the art would recognize that corrective filter 32 could include the ability to adjust the relative delays of the audio signals.

It could also be envisioned to embed the auto calibration process of surround

15   sound systems directly into so-called digital smart speakers (DSS) with a DSP and other supporting components implemented within the loudspeaker enclosure. Signals to these DSS loudspeakers could be analog or digital (or a combination of both analog and/or digital) and could convey audio information as well as loudspeaker identification information and electrical power. The user would simply connect any output of a

20   receiver to any speaker, letting the processors decode the information which is intended for that specific location. Since transfer rates of modern networks are at least in the

MHz range, technologies within the current art are fully adequate to support this level of functionality.

Accordingly, the spirit and scope of the present invention is to be limited only by the following claims.

5   What is claimed is: